Table 3 Summary of TSF and non-TSF-based traffic prediction

Ref.	ML Technique	Application *(approach)*	Dataset *(availability)*	Features	Output *(training)*	Evaluation Settings	Evaluation Results[a,b]
NBP [141]	Supervised: · MLP-NN *(offline)*	End-to-end path bandwidth availability prediction *(TSF)*	NSF TeraGrid dataset *(N/A)*	Max, Min, Avg load observed in past 10 s $\sim$ 30 s	Available bandwidth on a end-to-end path in future epoch	Number of features= 3 MLP-NN: · *(N/A)*	MSE $=$ 8%
Cortez et al. [104]	Supervised: · NNE trained with Rp *(offline)*	Link load and traffic volume prediction in ISP networks *(TSF)*	SNMP traffic data from 2 ISP nets, · traffic on a transatlantic link · aggregated traffic in the ISP backbone *(N/A)*	Traffic volume observed in past few minutes$\sim$several days	Expected traffic volume	Number of features= 6 $\sim$ 9 5 NNs NNE: · all SLPs for dataset1 · 1 hidden layer MLPs with 6 $\sim$ 8 neurons for dataset2	1h lookahead: · MAPE $=$ 1.43% $\sim$ 5.23% 1h $\sim$ 24h lookahead: · MAPE $=$ 6.34% $\sim$ 23.48%
Bermolen et al. [52]	Supervised: · SVR *(offline)*	Link load prediction in ISP networks *(TSF)*	Internet traffic collected at the POP of an ISP network *(N/A)*	Link load observed at τ time scale	Expected link load	Number of features= d samples with $d = 1..30$ Number of support vectors: · varies with d (e.g. $\sim$320 for $d = 10$)	RMSE < 2 for $\tau = 1ms$ and $d = 5$ · $\approx$ AR ·10% less than MA
Chabaa et al. [86]	Supervised: MLP-NN with different training algorithms (GD, CG, SS, LM, Rp) *(offline)*	Network traffic prediction *(TSF)*	1000 points dataset *(N/A)*	Past measurements	Expected traffic volume	Number of features *(N/A)* MLP-NN: · 1 hidden layer	LM: · RMSE= 0.0019 RPE $=$ 0.0230% Rp: · RMSE= 0.0031 RPE= 0.0371%
Zhu et al. [500]	Supervised: MLP-NN with PSO-ABC *(offline)*	Network traffic prediction *(TSF)*	2-week hourly traffic measurements *(N/A)*	N past days hourly traffic volume	Expected next-day hourly traffic volume	Number of features= 5 MLP-NN *(5, 11, 1)* PSO-ABC: · *30 particles of length=66*	MSE $=$ 0.006 on normalized data 50% less than BP
Li et al. [274]	Supervised: MLP-NN *(offline)*	Traffic volume prediction on an inter-DC link *(Regression)*	6-week inter-DC traffic dataset from Baidu · SNMP counters data collected every 30 s · Top-5 applications traffic data collected every 5 min *(N/A)*	Level-N wavelet transform used to extract time and frequency features from total and elephant traffic volumes time series	$k \times$ 30-s ahead expected traffic volume	Number of wavelets: · $N = 10$ Number of features= $k \times 120$ for $N = 10$ 1 hidden layer MLP-NN	RRMSE= 4% $\sim$ 10% for $k = 1 \sim 40$
Chen et al. [94]	Supervised: · KBR · LSTM-RNN *(offline)*	Inferring future traffic volume based on flow statistics *(regression)*	Network traffic volume and flow count collected every 5 min over a 24-week period *(public)*	Flow count	Expected traffic volume	Number of features: · 1 feature (past sample) LSTM-RNN: · *(N/A)*	RNN · MSE > 0.3 on normalized data · 0.05 higher than KBR · twice as much as RNN fed with traffic volume time series
Poupart et al. [365]	Supervised: · GPR · oBMM · MLP-NN *(offline)*	Early flow-size prediction and elephant flow detection *(classification)*	3 university and academic networks datasets with over three million flows each *(public)*	· source IP · destination IP · source port · destination port · protocol · server vs. client · size of 3 first packets	Flow size class; elephant vs. non-elephant	Number of features: · 7 features MLP-NN: · (106,60,40,1)	GPR: · TPR$>$ 80% · TNR$>$ 80% oBMM: · TPR and TNR $\approx$ 100% on one dataset · TPR$<$ 50% on other datasets MLP-NN: · TPR$>$ 80% · lowest TNR$<$ 80%

[a] Average values. Results vary according to experimental settings
[b] Accuracy metrics: mean square error (MSE), relative prediction error (RPE), mean absolute prediction error (MAPE), average root mean square error (RMSE)